

Configuration Manual

MSc Internship
Programme Name

Eze Kenneth C.
Student ID: X19131178

School of Computing
National College of Ireland

Supervisor: Mr. Niall Heffernan

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Eze Kenneth C.
Student ID: X19131178
Programme: Cybersecurity **Year:** 2019/2020
Module: Internship
Supervisor: Mr. Niall Heffernan
Submission Due Date: 17th August, 2020
Project Title: CYBERCRIME DETECTION IN COMMUNICATIONS: AN EXPERIMENTAL CASE OF CYBER SEXUAL HARRASMENT ACCURACY DETECTION ON TWITTER USING SUPERVISED LEARNING CLASSIFIERS

Word Count: **Page Count:**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:

Date:

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

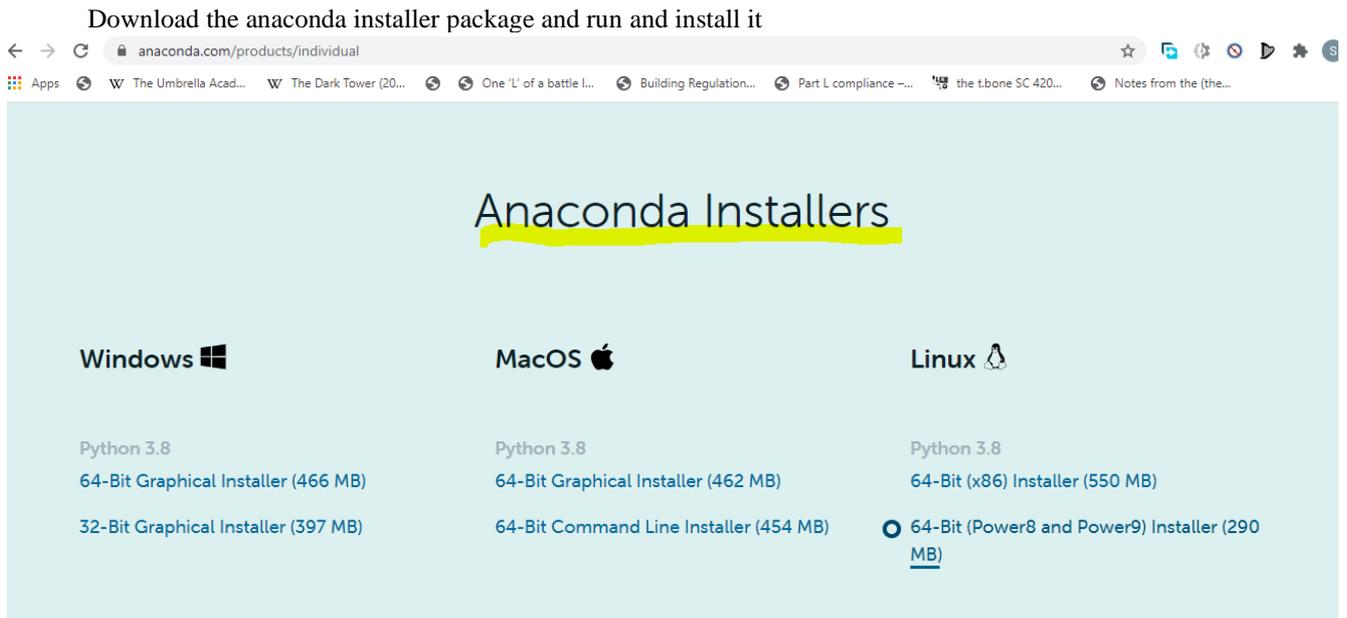
Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	

Penalty Applied (if applicable):	
----------------------------------	--

Eze Kenneth C.
Student ID: X19131178

1 Download the Anaconda Installer Package



1.1 Install Required libraries in the python command prompt

Syntax: pip install <library name>

- Pip install tweepy
- Textblob
- Vadersentiment
- Pandas
- Xgboost
- sklearn

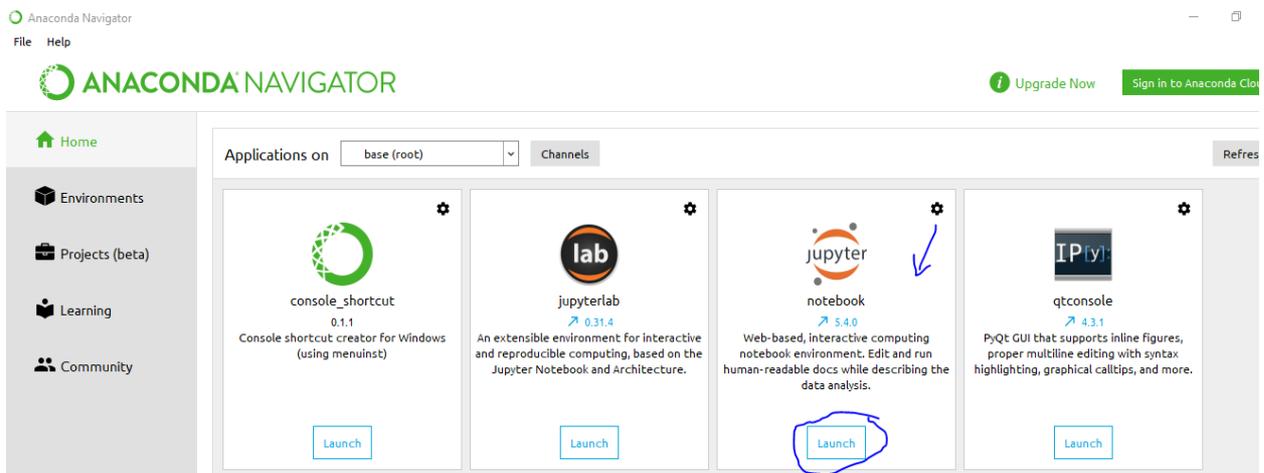
```

Anaconda Prompt (anaconda3)
(base) C:\Users\GERRARD>pip install tweepy
Collecting tweepy
  Downloading tweepy-3.9.0-py2.py3-none-any.whl (30 kB)
Collecting requests-oauthlib>=0.7.0
  Downloading requests_oauthlib-1.3.0-py2.py3-none-any.whl (23 kB)
Requirement already satisfied: requests[socks]>=2.11.1 in c:\users\gerrard\anaconda3\lib\site-packages (from tweepy) (2.22.0)
Requirement already satisfied: six>=1.10.0 in c:\users\gerrard\anaconda3\lib\site-packages (from tweepy) (1.14.0)
Collecting oauthlib=3.0.0
  Downloading oauthlib-3.1.0-py2.py3-none-any.whl (147 kB)
    | 147 kB 1.3 MB/s
Requirement already satisfied: certifi>=2017.4.17 in c:\users\gerrard\anaconda3\lib\site-packages (from requests[socks]>=2.11.1->tweepy) (2019.11.28)
Requirement already satisfied: idna<2.9,>=2.5 in c:\users\gerrard\anaconda3\lib\site-packages (from requests[socks]>=2.11.1->tweepy) (2.8)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in c:\users\gerrard\anaconda3\lib\site-packages (from requests[socks]>=2.11.1->tweepy) (3.0.4)
Requirement already satisfied: urllib3!=1.25.0,!<1.25.1,<1.26,>=1.21.1 in c:\users\gerrard\anaconda3\lib\site-packages (from requests[socks]>=2.11.1->tweepy) (1.25.8)
Requirement already satisfied: PySocks!=1.5.7,>=1.5.6; extra == "socks" in c:\users\gerrard\anaconda3\lib\site-packages (from requests[socks]>=2.11.1->tweepy) (1.7.1)
Installing collected packages: oauthlib, requests-oauthlib, tweepy
Successfully installed oauthlib-3.1.0 requests-oauthlib-1.3.0 tweepy-3.9.0

(base) C:\Users\GERRARD>

```

2 Now go into the Anaconda Navigation and launch the Jupyter Notebook



2.2 Locate the folder and click on the python program



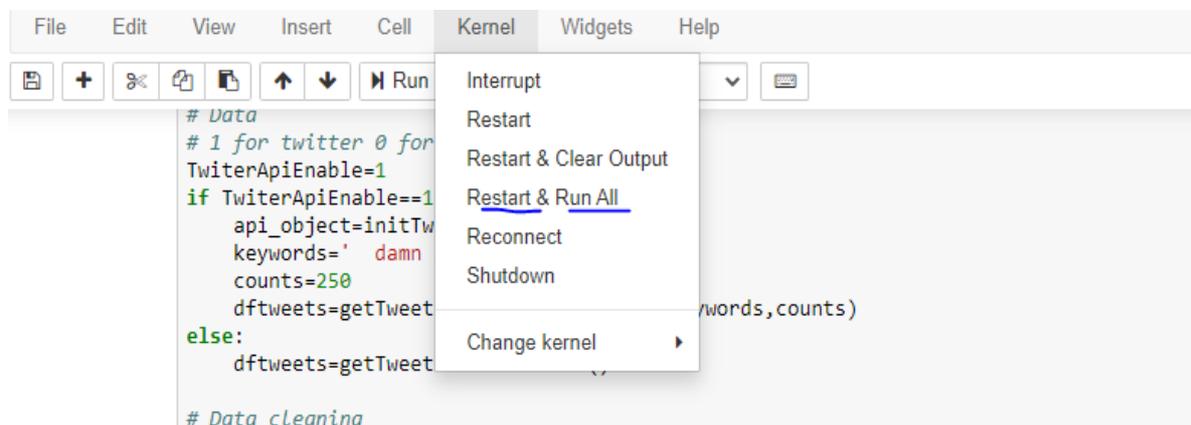
2.2 As we can see above, we have Kaggle csv file and python program. So, we can switch between the csv file in our testing and twitter api by enable either 1 for twitter api and 0.

Also, keywords for testing our accuracy

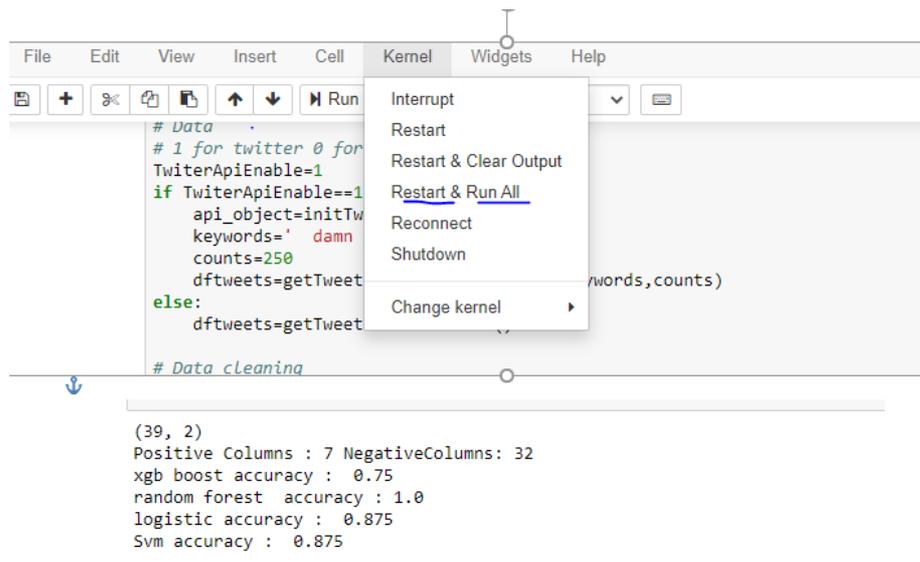
```
# program Stars
# Use this section to print result with data base
# Data
# 1 for twitter 0 for csv
TwiterApiEnable=1
if TwiterApiEnable==1:
    api_object=initTwitter()
    keywords=' damn stupid bitch like you'
    counts=250
    dftweets=getTweetsFromApi(api_object,keywords,counts)
else:
    dftweets=getTweetsFromDataBase()

# Data cleaning
dftweets=DataCleaning(dftweets)
# Pos Tagging
dftweets=PostaggingStemming(dftweets)
# model accuracy
model(dftweets)
```

2.2 Now go onto the kernel and launch the python program



2.3 Below we can see the output of the program for further analysis



The screenshot shows a Jupyter Notebook interface. The top menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. Below the menu is a toolbar with icons for saving, adding, deleting, and running cells. The main area contains a code cell with the following Python code:

```
# Data
# 1 for twitter 0 for
TwitterApiEnable=1
if TwitterApiEnable==1
    api_object=initTw
    keywords=' damn
    counts=250
    dftweets=getTweet
else:
    dftweets=getTweet
# Data cleaning
```

The output of the code cell is displayed below the code:

```
(39, 2)
Positive Columns : 7 NegativeColumns: 32
xgb boost accuracy : 0.75
random forest accuracy : 1.0
logistic accuracy : 0.875
Svm accuracy : 0.875
```

3 Mini Code Helper

The idea behind the code helper is to explain more to the reader some code functions they might find confusing in the code.

- **PostaggingStemming** - Using the nltk library we get the stop words here and clean the tweets. Also remove words which have characters less than 3 so that we only focus on words which are important
- **ParseTweet** - This function cleans the tweets and removes characters for ease of tokenization later.
- **DataCleaning** - This function is for cleaning the csv tweets.
- **getTweetsFromDataBase** - This function parses the tweets from the csv file and creates a data frame.
- **Model** - In this function we split the data in train and test and feed the train and test data in different models and get accuracy scores.